



## IMPRESS Code Review 2021: Submission on behalf of SumOfUs

### *An emergent gap in regulation*

We welcome the opportunity to respond to this important consultation. As the only body recognised by the Royal Charter on self-regulation of the press, IMPRESS is in a uniquely influential position to respond to evolving challenges posed by digital platforms. Against the backdrop of the UK government's white paper on online harms, this review of the Code presents an especially timely opportunity to establish new precedents for effective self-regulation of the press both in the UK and internationally.

SumOfUs is a global consumer advocacy organisation and online community with nearly 17 million members worldwide. For a number of years now, we have been at forefront of campaigns aimed at checking and reversing the spread of disinformation online. In November last year, we mobilised over 50 thousand activists within a week demanding that tech giants rework their algorithms to downgrade disinformation and hate speech.<sup>1</sup>

Though we were encouraged to see the UK government start to take similar action in its white paper on online harms, we are gravely concerned that news publishers will remain off the hook. The reality of the digital news ecosystem is that fake news and disinformation is not exclusively the product of major algorithms, foreign interference, or fringe and extremist groups. Both alternative (left and right) as well as mainstream press have been implicated in spreading or amplifying misinformation and existing regulatory frameworks and policies must be adapted and strengthened to confront this reality.

The current IMPRESS code review is an ideal starting point for addressing this emergent gap in policymaking and policy debate. There is one area in particular – user generated content – where we believe the Code should be adapted to reflect the changing dynamics of online news delivery and consumption.

The government's white paper on online harms clearly signals that the major intermediary platforms will be the exclusive targets of any legislation aimed at confronting the spread of disinformation and hate speech online, exempting publishers from any responsibility for user commentary on their own sites. The latter is likely to become more significant if any new legislation targeting the platforms is effective. We can and should expect racist and abusive commentary online to migrate and concentrate in those areas where it is less likely to be censored. This brings into sharp relief the role of press regulators in ensuring that publishers' platforms do not become a safe harbour for racism, misogyny and disinformation produced by users.

---

<sup>1</sup> See <https://www.sumofus.org/media/online-campaign-demanding-google-facebook-youtube-twitter--others-stop-pushing-harmful-content-garners-over-51k-supporters-in-less-than-a-week/>

At present, the Code’s preamble states that it is applicable “to all content and newsgathering activities for which publishers are responsible”. However, the particular and distinct responsibilities that should be assigned to publishers in respect of user generated content is not presently made clear either in the Code itself or the Regulatory Scheme.

Clearly, publishers should not be assigned the same degree of editorial responsibility for user generated content as for the original journalistic output that they produce. Such an equivalence would be wholly disproportionate and inconsistent with related precedents established in law in respect of a host platform’s liability for user generated content.

However, recent research has demonstrated the potential contribution that reader comments can make to the spread of misinformation and/or hate speech.<sup>2</sup> Reader commentary can also have an adverse influence over audience trust<sup>3</sup>.

We therefore recommend a provision be added to the Code which places **a specified duty of care on all publishers that facilitate reader comments**. This provision should emphasise the responsibility of such publishers **to adopt timely, consistent and transparent moderation policies and practices, based on a model standard developed by IMPRESS**.

We do not, however, consider that a complaints-based system of monitoring and enforcement would be appropriate for this aspect of the code, save for exceptional grounds and circumstances. We recommend instead that IMPRESS carries out **periodic review of a sample of reader comments to ensure adherence to the standards**. We believe that the appropriate sanction for publishers that fail to adhere to the standards should be a formal warning in the first instance, followed by an order to close all reader comments in the event of repeat failure.

#### *Proposed amendment*

The growing phenomenon of user participation in online news has posed myriad challenges for news publishers. Removing or restricting comment options for users risks diverting user participation to social media and ultimately impacting on financial bottom lines.<sup>4</sup> On the other hand, intervention in the form of active moderation can be both costly and risky, especially since the volume of commentary that any given article will attract is often difficult to predict.<sup>5</sup>

---

<sup>2</sup> Watson, B. R., Peng, Z., & Lewis, S. C. (2019). Who will intervene to save news comments? Deviance and social control in communities of news commenters. *New Media & Society*, 21(8), 1840-1858.

<sup>3</sup> Naab, T. K., Heinbach, D., Ziegele, M., & Grasberger, M. T. (2020). Comments and credibility: How critical user comments decrease perceived news article credibility. *Journalism Studies*, 21(6), 783-801.

<sup>4</sup> See <https://whatsnewinpublishing.com/publishers-that-closed-their-comments-sections-made-a-colossal-mistake/>

<sup>5</sup> Ambroselli, C., Risch, J., Krestel, R., & Loos, A. (2018, June). Prediction for the newsroom: Which articles will get the most comments?. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)* (pp. 193-199).

Little wonder then, that recent research overwhelmingly demonstrates a lack of consistency and transparency in comment moderation practices. A 2019 study of moderation practices by online German news sites found that even within a given news organisation, there is often significant variance in approaches between moderators.<sup>6</sup> As a result, decisions about which comments to censor – whether taken by human editors or artificial intelligence – are often ineffective or biased. This can in turn have a potentially damaging impact on reader trust. A recent study showed that news sites which employed automated moderation were more trusted compared to those that used human agents.<sup>7</sup>

Ineffective and inconsistent filtering practices can also exacerbate the degree to which reader comments can contribute to the spread of fake news, disinformation and hate speech. A study of reader comments on the *Guardian* website in 2018 found that female and BAME journalists were disproportionately subject to abuse by reader comments.<sup>8</sup>

But often harder to detect by comment moderators is the phenomenon of ‘dark participation’ which has become synonymous with disruptive tactics used by lobbyists, pressure groups and political trolls to manipulate online conversation. Reader comments provide an opportunity for such actors to reach a potentially much larger audience than might be possible on social media.<sup>9</sup>

There is, however, little evidence to suggest that reader comments have a significant influence over public conversation, and there is a risk of regulatory overreach, especially given the added cost and resource pressure that comment moderation entails. Whilst there is clearly a need for regulatory catch up, we recommend that any addition to the code in this area is relatively light touch.

Specifically, we are calling for a clause to be added to the code which imposes a duty of care on publishers that facilitate reader comments. This duty of care should specify the responsibility of such publishers to filter out fake news, disinformation, hate speech and otherwise abusive and disruptive comment behaviour, and for this to be done in a manner that is timely, consistent and transparent.

We further recommend emphasis to be placed on articles that are focused on political controversies or social issues, as these are more likely to attract potentially harmful reader participation. Equally, emphasis should be placed on comment moderation within 24 hours of publication. This is the period during which any given news article can be reasonably considered ‘live’ and the period in which its audience reach is relatively concentrated. Recent research also suggests that majority of reader comments are posted within a few hours of publication.<sup>10</sup>

---

<sup>6</sup> Frischlich, L., Boberg, S., & Quandt, T. (2019). Comment sections as targets of dark participation? Journalists’ evaluation and moderation of deviant user comments. *Journalism Studies*, 20(14), 2014-2033.

<sup>7</sup> Wang, S. (2021). Moderating Uncivil User Comments by Humans or Machines? The Effects of Moderation Agent on Perceptions of Bias and Credibility in News Content. *Digital Journalism*, 9(1), 64-83.

<sup>8</sup> Gardiner, B. (2018). “‘It’s a terrible way to go to work:’ what 70 million readers’ comments on the Guardian revealed about hostility to women and minorities online.” *Feminist Media Studies* 18.4 (2018): 592-608.

<sup>9</sup> Quandt, T. (2018). Dark participation. *Media and Communication*, 6(4), 36-48.

<sup>10</sup> He, L., Han, C., Mukherjee, A., Obradovic, Z., & Dragut, E. (2020). On the dynamics of user engagement in news comment media. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(1), e1342.

In regard to monitoring and enforcement, we believe this area also calls for a distinct approach and that a system based exclusively on complaints would not be appropriate. Admitting complaints based on individual reader comments could give rise to a volume of complaints that is disproportionate to the likely impact of any such comment, especially in the case of publishers with a relatively small audience reach. Under a complaints-based system of monitoring and enforcement, smaller publishers could face intense additional resource pressure resulting from complaints handling on top of what is required for the moderation of complaints. We believe a more proportionate and effective approach to monitoring and enforcement would be for IMPRESS to periodically review a sample of reader comments from publishers with a focus on those with an online reach of over one million unique visitors per month. This periodic and relatively cursory review could then trigger more in-depth reviews of particular publishers where questionable practices are identified.

As well as being proportionate and effective, such an approach has the advantage of not placing an undue burden on the regulator. There may, however, be scope for admitting individual complaints in particular circumstances, such as when a relatively large publisher has failed to remove comments that are liable to cause extreme offence. But the defined grounds for any such complaints should make clear that they are exceptional.

In regard to sanction and remedy, we believe that a relatively simple approach could be adopted whereby publishers found to have failed in their duty of care are given a formal warning in the first instance. Any subsequent or repeat failure should then trigger an order to stop facilitating reader comments or any form of user generated content.

In summary, we believe there is a need for the Code to be adapted to make clear the scope of responsibilities that publishers have in respect of user generated content on their platforms. It is equally imperative, however, that any additional regulation so entailed must be proportional to the likely reach and impact of *individual* reader comments, as well as take account of the additional resource pressures that could adversely impact on smaller independent publishers.